ELSEVIER

European Journal of Operational Research 138 (2002) 191–211

EUROPEAN JOURNAL OF OPERATIONAL RESEARCH

www.elsevier.com/locate/dsw

Interfaces with Other Disciplines

Bayesian neural network learning for repeat purchase modelling in direct marketing

Bart Baesens [a], Stijn Viaene [a], Dirk Van den Poel [b,*], Jan Vanthienen [a], Guido Dedene [a]

[a] *K.U.Leuven, Department of Applied Economic Sciences, Naamsestraat 69, B-3000 Leuven, Belgium*
[b] *Ghent University, Department of Marketing, Hoveniersberg 24, B-9000 Ghent, Belgium*

Received 7 November 2000; accepted 14 March 2001

Abstract

We focus on purchase incidence modelling for a European direct mail company. Response models based on statistical and neural network techniques are contrasted. The evidence framework of MacKay is used as an example implementation of Bayesian neural network learning, a method that is fairly robust with respect to problems typically encountered when implementing neural networks. The automatic relevance determination (ARD) method, an integrated feature of this framework, allows us to assess the relative importance of the inputs. The basic response models use operationalisations of the traditionally discussed Recency, Frequency and Monetary (RFM) predictor categories. In a second experiment, the RFM response framework is enriched by the inclusion of other (non-RFM) customer profiling predictors. We contribute to the literature by providing experimental evidence that: (1) Bayesian neural networks offer a viable alternative for purchase incidence modelling; (2) a combined use of all three RFM predictor categories is advocated by the ARD method; (3) the inclusion of non-RFM variables allows to significantly augment the predictive power of the constructed RFM classifiers; (4) this rise is mainly attributed to the inclusion of customer/company interaction variables and a variable measuring whether a customer uses the credit facilities of the direct mailing company.

Keywords: Neural networks; Marketing; Bayesian learning; Response modelling; Input ranking

1. Introduction

It is well established in the literature that customer retention is at least as important as customer acquisition in the current context of competitive markets, not in the least for (direct) mail-order companies. Mail-order companies typically are in the business of sending out catalogues to a selected number of prospective buyers. The

* Corresponding author. Tel.: +32-9-264-8980; fax: +32-9-264-4279.

E-mail addresses: bart.baesens @ econ.kuleuven.ac.be (B. Baesens), stijn.viaene@ econ.kuleuven.ac.be (S. Viaene), dirk.vandenpoel@rug.ac.be (D. Van den Poel), jan.vanthienen@econ.kuleuven.ac.be (J. Vanthienen), guido.dedene@econ.kuleuven.ac.be (G. Dedene).

0377-2217/02/$ - see front matter
PII: S0377-2217(01)00129-1

selection of whom to include in the mailing list rests on an assessment of the individual's propensity to buy. The prospects or customers to be mailed are typically selected following the results of statistical models including behavioural, demographic and other customer profiling predictors in order to optimise the prospective buyer response rate. Commonly used target variables for these mailing response models are purchase incidence, purchase amount and interpurchase time. In this paper, we focus on the purchase incidence, i.e. the issue whether or not a purchase is made from any product category offered by the direct mail company.

Conceptually, the purchase incidence response modelling issue reduces to the general problem category of binary classification: repurchase or not. Among the traditional (statistical) techniques that have been widely used are logistic regression, linear and quadratic discriminant analysis models. However, their pre-determined functional form and restrictive (often unfounded) model assumptions limit their usefulness [4,58]. In this paper, we use neural networks (NNs) for response modelling. Their universal approximation property makes them a very interesting alternative for pattern recognition purposes. Unfortunately, many practical problems still remain when implementing NNs, e.g. *How to choose the appropriate number of hidden neurons? What is the impact of the initial weight choice? How to set the weight decay parameter? How to avoid the network from fitting noise in the training data?* These issues are often dealt with in an ad hoc way [3]. Nevertheless, they are crucial to the success of the NN implementation. A Bayesian learning paradigm has been suggested as a way to deal with these problems during NN training [4,33,34,42]. Here, all prior assumptions are made explicit and the weights and hyperparameters are determined by applying Bayes' theorem to map the prior assumptions into posterior knowledge after having observed the training data. In this paper, we use the evidence framework of MacKay as an example implementation of Bayesian learning [33–36]. An interesting additional feature of this framework is the automatic relevance determination (ARD) method which allows us to assess the relative importance of the various inputs by adding weight regularisation terms to the objective function. In this paper, it is shown that training NNs using the evidence framework (with the ARD extension) is an effective and viable alternative for the response modelling case at hand when compared to the three benchmark statistical techniques mentioned above.

The empirical study consists of two subexperiments. Initially, only standard Recency, Frequency and Monetary (RFM) predictor categories will underly the purchase incidence model. This choice is motivated by the fact that most previous research cites them as being most important and because they are internally available at very low cost [1,15,29]. It is shown for this case that, from a predictive performance perspective, Bayesian NNs are statistically superior when compared to logistic regression, linear and quadratic discriminant analysis classifiers. Predictive performance is quantified by means of the percentage correctly classified (PCC) and the area under the receiver operating characteristic curve (AUROC). The latter basically illustrates the behaviour of a classifier without regard to class distribution or error cost, so it effectively decouples classification performance from these factors [20,59,60]. The ARD method is used to shed light upon the relative importance of all variables operationalising the RFM response model. In a second experiment, the response model is extended with other potentially interesting customer profiling variables. It is illustrated that the Bayesian NNs still perform significantly better than the three statistical classifiers. Again, the relative importance of the inputs is assessed using the ARD method.

This paper is organised as follows. In Section 2 we provide a concise overview of response modelling issues in the context of direct marketing. Section 3 discusses the theoretical underpinnings of NNs for pattern recognition purposes. The Bayesian evidence framework for classification is presented in Section 4. Section 5 presents the ARD extension of the evidence framework. The design of the study, including data set description, experimental setup and used performance criteria are presented in Section 6. Results and discussion of